



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

LINEAR ALGEBRA
AND ITS
APPLICATIONS

Linear Algebra and its Applications 412 (2006) 39–74

www.elsevier.com/locate/laa

On approximately simultaneously diagonalizable matrices

K.C. O'Meara, C. Vinsonhaler *

Department of Mathematics, University of Connecticut, Storrs, CT 06269-3009, United States

Received 28 October 2004; accepted 19 April 2005

Available online 30 August 2005

Submitted by S. Kirkland

Dedicated to the memory of our esteemed colleague Kostia Beidar

Abstract

A collection A_1, A_2, \dots, A_k of $n \times n$ matrices over the complex numbers \mathbb{C} has the ASD property if the matrices can be perturbed by an arbitrarily small amount so that they become simultaneously diagonalizable. Such a collection must perforce be commuting. We show by a direct matrix proof that the ASD property holds for three commuting matrices when one of them is 2-regular (dimension of eigenspaces is at most 2). Corollaries include results of Gerstenhaber and Neubauer–Sethuraman on bounds for the dimension of the algebra generated by A_1, A_2, \dots, A_k . Even when the ASD property fails, our techniques can produce a good bound on the dimension of this subalgebra. For example, we establish $\dim \mathbb{C}[A_1, \dots, A_k] \leq 5n/4$ for commuting matrices A_1, \dots, A_k when one of them is 2-regular. This bound is sharp. One offshoot of our work is the introduction of a new canonical form, the H-form, for matrices over an algebraically closed field. The H-form of a matrix is a sparse “Jordan like” upper triangular matrix which allows us to assume that any commuting matrices are also upper triangular. (The Jordan form itself does not accommodate this.)

© 2005 Elsevier Inc. All rights reserved.

AMS classification: 15A21; 15A18; 15A60*Keywords:* Matrices; Canonical form; Simultaneous diagonalization

* Corresponding author.

E-mail addresses: staf198@ext.canterbury.ac.nz (K.C. O'Meara), vinsonhaler@math.uconn.edu (C. Vinsonhaler).

0. Introduction

In a recent study of phylogenetic invariants in biomathematics [1], the following question arose: Given A_1, A_2, \dots, A_k commuting $n \times n$ matrices over the complex numbers \mathbb{C} , can the matrices be perturbed by an arbitrarily small amount so that they become simultaneously diagonalizable? More specifically, given $\epsilon > 0$, are there $n \times n$ matrices E_i , with $\|E_i\| < \epsilon$ and an invertible $n \times n$ matrix C such that $C^{-1}(A_i + E_i)C$ is diagonal for $i = 1, 2, \dots, k$? Here the norm $\|X\|$ of a matrix X can be the square root of the sum of the absolute values squared of the entries, the maximum of the absolute values of the entries, or any other reasonable (equivalent) choice satisfying $\|XY\| \leq c\|X\|\|Y\|$ (c a constant) and $\|X + Y\| \leq \|X\| + \|Y\|$. Any list of matrices with the property in question will be called *approximately simultaneously diagonalizable*, abbreviated ASD. Note that we do not assume commutativity for this definition. However, as we show below (Proposition 2.2), the ASD property implies the commutativity of the matrices in the list.

The ASD property appears to have been studied only tangentially in the literature, mainly in connection with some problems in algebraic geometry. For instance, several authors (see [4,6,7,12]) have studied the variety $\mathcal{C}(d, n)$ of d -tuples of commuting $n \times n$ matrices over an algebraically closed field, particularly for $d = 3$ and small n . When this variety is known to be irreducible (nonempty open sets are dense), one gets the ASD property for d commuting matrices as a corollary from the observation that the d -tuples in which the first matrix has n distinct eigenvalues form an open set. (This argument can sometimes also be applied to suitable subvarieties of $\mathcal{C}(d, n)$.) However, the ASD property itself appears to be a weaker condition than the irreducibility of $\mathcal{C}(d, n)$ (or of various subvarieties) and, moreover, not much is known about the latter. For example, even the irreducibility of $\mathcal{C}(3, 5)$ is open whereas one can establish, via our techniques, that any three commuting 5×5 matrices have the ASD property (see Remark 6.2). We feel that the ASD and related concepts are interesting enough on their own to warrant a study by purely matrix-theoretic methods. This paper is a step in that direction.

The answer to the ASD question is negative in general (see Example 2.7). However, Motzkin and Taussky, in a theorem only incidental to the main thrust of their 1955 paper [10], proved that any two commuting matrices over \mathbb{C} are ASD. The ASD property also holds for three commuting matrices when one of them is 2-regular (definition in Section 1). This can be deduced from results in the 1999 Neubauer and Sethuraman paper [12], which employs methods of algebraic geometry. We will give a purely matrix-theoretic proof of the result (Theorem 6.1) in Sections 6 and 7.

As we show, the ASD property is tied to a more classical problem, that of bounding the dimension over \mathbb{C} of $\mathbb{C}[A_1, A_2, \dots, A_k]$, the subalgebra (with identity) of the $n \times n$ complex matrices generated by commuting A_1, A_2, \dots, A_k . Here the definitive result was proved by Gerstenhaber [4] in 1961: Any two commuting $n \times n$ matrices over an algebraically closed field generate a subalgebra of dimension no greater than n . (A number of authors have expanded and refined Gerstenhaber's result. See [2,6,9,11–13].) In 1999, Neubauer and Sethuraman [12] showed that this bound on

dimension still holds for three commuting matrices when one of them is 2-regular. We derive both the Gerstenhaber and the Neubauer and Sethuraman results in the complex case as corollaries to our ASD results, via the connection that ASD $n \times n$ matrices can generate a subalgebra of dimension at most n (Theorem 2.5). As a bonus, even when the ASD property fails, our techniques can still sometimes yield a good bound on the dimension of this subalgebra. For example, we establish in Theorem 8.2 the seemingly new result that $\dim \mathbb{C}[A_1, \dots, A_k] \leq 5n/4$ for commuting matrices A_1, \dots, A_k when one of them is 2-regular. (This bound is sharp.)

One possibly interesting offshoot of our work is the introduction of a new canonical form, the H-form, for matrices over an algebraically closed field. (“H” stands for “Husky”—in recognition of the University of Connecticut connection.) The H-form of a matrix is a sparse “Jordan like” upper triangular matrix which allows us to assume that any other commuting matrices are also upper triangular. (The Jordan form itself does not accommodate this.) Not only is the H-form a useful tool for constructing nice perturbations of commuting matrices, it also provides natural candidates for $n \times n$ commuting matrices which generate larger than normal commutative subalgebras. We plan to present examples of such subalgebras in a later paper.

1. Preliminaries

Every matrix over \mathbb{C} is similar to a matrix in Jordan form, $J = \text{diag}(J_1, \dots, J_s)$, a block diagonal matrix where each block has the form

$$J_i = \begin{bmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & \lambda_i & 1 \\ & & & & \lambda_i \end{bmatrix}.$$

If J has only one eigenvalue ($\lambda_i = \lambda_j$ for all i, j), then we say J has *Jordan structure* (m_1, m_2, \dots, m_s) , where $m_i \geq m_{i+1}$ and J_i is $m_i \times m_i$.

A matrix is called *l-regular* if in its Jordan form at most l blocks share the same eigenvalue. Thus, a 1-regular matrix (also called *regular* or *nonderogatory*) is one for which each eigenspace has dimension 1. Later we will work primarily with 2-regular matrices, so the dimension of each eigenspace is at most two. The norm $\|A\|$ of a square matrix $A = (a_{ij})$ is given by $\|A\|^2 = \sum |a_{ij}|^2$. Any “reasonable” norm would serve our purposes equally well, for example $\|A\| = \max\{|a_{ij}|\}$. By $n \times n$ matrices B_1, \dots, B_k (over a field) being *simultaneously diagonalizable* we mean, of course, that there exists an invertible matrix C such that $C^{-1}B_1C, \dots, C^{-1}B_kC$ are diagonal matrices. It is the “approximate” version of this that is the focus of our paper.

Definition 1.1. We say that complex $n \times n$ matrices A_1, A_2, \dots, A_k are *approximately simultaneously diagonalizable* (abbreviated ASD) if for any $\epsilon > 0$, there exist matrices B_1, B_2, \dots, B_k which are simultaneously diagonalizable and satisfy

$$\|B_i - A_i\| < \epsilon \quad \text{for } i = 1, \dots, k.$$

2. Some results on ASD matrices

One of the earliest results on the ASD property is the following 1956 theorem of Motzkin and Taussky [10, Theorem 5].

Theorem 2.1. *Every pair of complex commuting $n \times n$ matrices has the ASD property.*

The proof is surprisingly straightforward, but depends very much on there being only two commuting matrices. We will outline the proof in Section 3 (following 3.4). The next proposition shows that matrices with the ASD property must necessarily commute.

Proposition 2.2. *If A_1, A_2, \dots, A_k are ASD, then $A_i A_j = A_j A_i$ for all i, j .*

Proof. Suppose, for example, that A_1 and A_2 do not commute. Since the commutator mapping $(X, Y) \rightarrow [X, Y] = XY - YX$ from $M_n(\mathbb{C}) \times M_n(\mathbb{C})$ to $M_n(\mathbb{C})$ is continuous and $[A_1, A_2] \neq 0$, there exists $\epsilon > 0$ such that $[B_1, B_2] \neq 0$ for all B_1, B_2 with $\|B_i - A_i\| < \epsilon$. In this case B_1 and B_2 do not commute so they cannot be simultaneously diagonalizable. This contradicts the ASD hypothesis, completing the proof. \square

There is one important consequence of the ASD property for a collection of $n \times n$ matrices that seems to have gone unnoticed to this point. Namely, the subalgebra these matrices generate can have dimension at most n . We aim now to establish this property.

Proposition 2.3. *If A_1, A_2, \dots, A_k are linearly independent in $M_n(\mathbb{C})$, then there exists $\epsilon > 0$ such that if B_i satisfies $\|B_i - A_i\| < \epsilon$ for $i = 1, 2, \dots, k$, then B_1, B_2, \dots, B_k are also linearly independent.*

Proof. It is enough to establish the result for vectors, in fact for m linearly independent vectors v_1, \dots, v_m in \mathbb{C}^m (after expanding the original set to a basis). Let M be the $m \times m$ matrix with v_1, \dots, v_m as its columns. Since $\det : M_m(\mathbb{C}) \rightarrow \mathbb{C}$ is a continuous function, and $\det M \neq 0$, there is an open neighborhood \mathcal{N} of M such that $\det X \neq 0$ for all $X \in \mathcal{N}$. Since any such X has independent columns, the result follows. \square

Lemma 2.4. *Let \mathcal{A} be a commutative subalgebra (with identity) of $M_n(\mathbb{C})$ and suppose A_1, \dots, A_k generate \mathcal{A} as an algebra. If A_1, \dots, A_k are ASD, then so also is any finite set of matrices in \mathcal{A} .*

Proof. There exists a \mathbb{C} -vector space basis for \mathcal{A} of monomials M_1, \dots, M_r in the A_i , say of degree at most d . We can assume M_1 has degree 0 ($M_1 = I$) and the other M_j have positive degree. Given $\epsilon > 0$, let $b = \max\{\|A_1\|, \dots, \|A_k\|, 1\}$ and $\epsilon' = \epsilon/2^d b^{d-1}$. Suppose $A_1 + E_1, \dots, A_k + E_k$ are simultaneously diagonalizable approximations of A_1, A_2, \dots, A_k with $\|E_i\| < \epsilon'$ for $i = 1, 2, \dots, k$. Substitute $A_i + E_i$ for A_i in the monomials M_j to obtain monomials M'_j in the $A_i + E_i$. We can expand M'_j as a sum of the monomial M_j and monomial terms involving error terms E_i as well as the original matrices A_i . Each term involves at most $d - 1$ matrices A_i and there are $2^d - 1$ such terms. Thus, $\|M'_j - M_j\| < 2^d b^{d-1} \epsilon' = \epsilon$. That is, the basis M_1, \dots, M_r can be approximated by simultaneously diagonalizable matrices.

Now let X_1, \dots, X_s be any finite subset of \mathcal{A} . For $i = 1, \dots, s$, write $X_i = \sum_{j=1}^r c_{ij} M_j$ and let $c = \max_{i,j} \{|c_{ij}|\}$. Given $\epsilon > 0$, let M'_1, \dots, M'_r be simultaneously diagonalizable ϵ -approximations of M_1, \dots, M_r . Set $X'_i = \sum_{j=1}^r c_{ij} M'_j$. Then X'_1, \dots, X'_s are simultaneously diagonalizable and $\|X'_i - X_i\| = \|\sum_{j=1}^r c_{ij} (M'_j - M_j)\| \leq \sum_{j=1}^r |c_{ij}| \|M'_j - M_j\| < rc\epsilon$. Hence X_1, \dots, X_s can be approximated by simultaneously diagonalizable matrices, as asserted. \square

Theorem 2.5. *If a commutative subalgebra \mathcal{A} of $M_n(\mathbb{C})$ has a finite set of generators that can be approximated by simultaneously diagonalizable matrices, then $\dim \mathcal{A} \leq n$.*

Proof. Let $r = \dim \mathcal{A}$ and let $\{B_1, \dots, B_r\}$ be a vector space basis for \mathcal{A} . By Lemma 2.4, B_1, \dots, B_r can be approximated by simultaneously diagonalizable matrices B'_1, \dots, B'_r . Moreover, by Proposition 2.3, we can arrange for B'_1, \dots, B'_r to be linearly independent. Let C be an invertible matrix such that for each i , $C^{-1} B'_i C = D_i$, a diagonal matrix. Since D_1, \dots, D_r are linearly independent members of the n -dimensional space of diagonal matrices, $r = \dim \mathcal{A} \leq n$. \square

As a corollary, we obtain a novel proof of a special case of Gerstenhaber's 1961 theorem [4, Theorem 2].

Corollary 2.6 (Gerstenhaber). *Every 2-generator commutative subalgebra of $M_n(\mathbb{C})$ has dimension at most n .*

Proof. Suppose the subalgebra \mathcal{A} of $M_n(\mathbb{C})$ is generated by commuting matrices A, B . By Theorem 2.1, A and B are ASD. Therefore by Theorem 2.5, $\dim \mathcal{A} \leq n$. \square

Allman and Rhodes have observed in [1, Lemma 9], that for $n = 1, 2, 3, 4$ any $n - 1$ commuting $n \times n$ complex matrices have the ASD property. To our knowledge,

no failures of the ASD property for commuting matrices have been explicitly recorded in the literature. We give some examples below.

Example 2.7. For each $n \geq 4$, there exist n commuting $n \times n$ complex matrices A_1, \dots, A_n which fail the ASD property.

Proof. Firstly we consider the case $n = 4$, and let $A_1 = e_{13}$, $A_2 = e_{14}$, $A_3 = e_{23}$, $A_4 = e_{24}$. (Here e_{ij} denotes the matrix unit with a 1 in the (i, j) position and zeroes elsewhere.) Notice that all the products $A_i A_j$ are zero, whence A_1, \dots, A_4 generate the commutative subalgebra (with identity)

$$\begin{aligned} \mathcal{A} &= \text{set of scalar matrices} + \text{linear span of } A_1, \dots, A_4 \\ &= \left\{ \begin{bmatrix} a & 0 & b & c \\ 0 & a & d & e \\ 0 & 0 & a & 0 \\ 0 & 0 & 0 & a \end{bmatrix} : a, b, c, d, e \in \mathbb{C} \right\}. \end{aligned}$$

Since $\dim \mathcal{A} = 5 > 4 = n$, by Theorem 2.5, A_1, \dots, A_4 fail the ASD property.

In general, for $n \geq 4$ we proceed in an entirely similar fashion by selecting any n matrix units e_{ij} from the “upper right hand corner”, that is, with $1 \leq i \leq \frac{n}{2}$ and $\frac{n}{2} < j \leq n$. (This, note, is not possible when $n < 4$.) These matrix units are independent and all their products are zero, so they generate a commutative subalgebra \mathcal{A} (with identity) with $\dim \mathcal{A} = n + 1 > n$. Again, Theorem 2.5 says that the chosen n matrix units fail the ASD property. \square

3. Splittings induced by epsilon changes

Our methods (later) for establishing the ASD and related properties rely on a standard block splitting of commuting matrices, which we use to harness induction arguments. The splitting is recorded in the following proposition, which also shows how the ASD problem reduces to commuting nilpotent matrices.

Proposition 3.1. Suppose A_1, \dots, A_k are commuting $n \times n$ matrices over an algebraically closed field F . Then there exists an invertible matrix C such that $C^{-1}A_1C, \dots, C^{-1}A_kC$ are block diagonal matrices with matching block structures and each diagonal block has only a single eigenvalue (ignoring multiplicities). That is, there is a partition $n = n_1 + \dots + n_r$ of n such that

$$C^{-1}A_iC = \begin{bmatrix} B_{i1} & & & \\ & B_{i2} & & \\ & & \ddots & \\ & & & B_{ir} \end{bmatrix},$$

where each B_{ij} is an $n_j \times n_j$ matrix having only a single eigenvalue for $i = 1, \dots, k$ and $j = 1, \dots, r$. Moreover, if $B_{1j}, B_{2j}, \dots, B_{kj}$ are ASD for $j = 1, \dots, r$, then A_1, A_2, \dots, A_k are ASD.

Remark 3.2. Each $B_{ij} = \lambda_{ij}I + N_{ij}$ for some scalar matrix $\lambda_{ij}I$ and nilpotent matrix N_{ij} . Clearly B_{1j}, \dots, B_{kj} are ASD if and only if N_{1j}, \dots, N_{kj} are ASD.

Proof. Let $\lambda_1, \dots, \lambda_r$ be the distinct eigenvalues of A_1 , and let $p(\lambda) = (\lambda - \lambda_1)^{n_1} \dots (\lambda - \lambda_r)^{n_r}$ be its characteristic polynomial. Let $V = F^n$. Then, regarding A_1 as a linear transformation on V ,

$$V = \ker(\lambda_1 I - A_1)^{n_1} \oplus \dots \oplus \ker(\lambda_r I - A_1)^{n_r}.$$

Since A_1, \dots, A_k commute, each $\ker(\lambda_i I - A_1)^{n_i}$ is invariant under A_1, \dots, A_k . Therefore, by choosing an invertible matrix C whose columns are (groups of) basis elements for the various $\ker(\lambda_i I - A_1)^{n_i}$, we get

$$C^{-1}A_iC = \begin{bmatrix} B_{i1} & & & \\ & B_{i2} & & \\ & & \ddots & \\ & & & B_{ir} \end{bmatrix} \quad \text{for } i = 1, \dots, k,$$

where B_{ij} is an $n_j \times n_j$ matrix. Moreover, each B_{1j} has only a single eigenvalue (namely λ_j). For fixed j , $B_{1j}, B_{2j}, \dots, B_{kj}$ must commute, and if one of them has two distinct eigenvalues we can repeat the splitting on these blocks, and so on. Eventually (or by induction), we achieve a splitting in which each B_{ij} has only a single eigenvalue. (We do not exclude the possibility that some B_{ij} and B_{im} could share the same eigenvalue for $j \neq m$.)

Now suppose $B_{1j}, B_{2j}, \dots, B_{kj}$ are ASD for $j = 1, \dots, r$. Let $\epsilon > 0$. For $j = 1, \dots, r$ choose simultaneously diagonalizable matrices $B'_{1j}, B'_{2j}, \dots, B'_{kj}$ such that

$$\|B'_{ij} - B_{ij}\| < \frac{\epsilon}{n\|C\| \cdot \|C^{-1}\|}$$

for all i, j . Let

$$B'_i = \begin{bmatrix} B'_{i1} & & & \\ & B'_{i2} & & \\ & & \ddots & \\ & & & B'_{ir} \end{bmatrix}$$

and let $A'_i = CB'_iC^{-1}$ for $i = 1, \dots, k$. Then B'_1, \dots, B'_k are simultaneously diagonalizable, whence so are A'_1, \dots, A'_k . Now

$$\begin{aligned} \|A'_i - A_i\| &= \|CB'_iC^{-1} - C(C^{-1}A_iC)C^{-1}\| \\ &= \|C(B'_i - C^{-1}A_iC)C^{-1}\| \\ &\leq \|C\| \cdot \|C^{-1}\| \cdot \|B'_i - C^{-1}A_iC\| \\ &\leq \|C\| \cdot \|C^{-1}\| \sum_{j=1}^r \|B'_{ij} - B_{ij}\| \\ &< \|C\| \cdot \|C^{-1}\| \cdot \frac{r\epsilon}{n\|C\| \cdot \|C^{-1}\|} \\ &\leq \epsilon. \end{aligned}$$

This demonstrates that A_1, \dots, A_k are ASD. \square

Notice that the splitting in Proposition 3.1 will be nontrivial ($r > 1$) if one of A_1, \dots, A_k has at least two distinct eigenvalues. When this does occur, it kicks in a natural induction on the smaller sized block diagonal matrices for establishing the ASD and related properties.

Let us say that an ϵ -perturbation $\bar{B} = B + E$ of an $n \times n$ matrix B (that is, $\|E\| < \epsilon$) is *k-correctable* if given any finite collection $B_1 = B, B_2, \dots, B_k$ of k commuting matrices, there are ϵ -perturbations $\bar{B}_2, \dots, \bar{B}_k$ of B_2, \dots, B_k such that $\bar{B}_1, \bar{B}_2, \dots, \bar{B}_k$ still commute. Correctable perturbations are not easy to spot. However, a key strategy later for establishing the ASD property for certain commuting nilpotent matrices A_1, \dots, A_k is to make a *k-correctable* ϵ -perturbation of A_1 which produces two eigenvalues 0 and ϵ . We then split commuting perturbed matrices $\bar{A}_1, \dots, \bar{A}_k$ using Proposition 3.1, and repeat the argument (inductively) on smaller nilpotent matrices. The following proposition warns us, however, that we cannot expect to always have the ϵ -eigenspaces one-dimensional at each step. We denote the centralizer of a square matrix A by $\mathcal{C}(A)$.

Proposition 3.3. *Suppose A is an $n \times n$ nilpotent matrix with $\dim \mathcal{C}(A) > n$. Then A cannot be perturbed to a diagonalizable matrix by a series of n arbitrarily small 2-correctable perturbations which introduce one new eigenvalue (of multiplicity one) at each stage.*

Proof. Let $\epsilon > 0$. Suppose B_1, \dots, B_n are successive 2-correctable ϵ -perturbations of A which introduce one new eigenvalue at each step. Then B_n is diagonalizable with n distinct eigenvalues and so $\dim \mathcal{C}(B_n) = n$. Let $\{C_1, \dots, C_m\}$ be a basis for $\mathcal{C}(A)$. By Proposition 2.3 we can arrange the choice of ϵ so that any series of up to n ϵ -perturbations of C_1, \dots, C_m will preserve their linear independence. Since B_1 is a 2-correctable perturbation of A , there are ϵ -perturbations $\bar{C}_1, \dots, \bar{C}_m$ of C_1, \dots, C_m such that $\bar{C}_1, \dots, \bar{C}_m$ centralize B_1 . Hence $\dim \mathcal{C}(B_1) \geq \dim \mathcal{C}(A)$. Repeating this argument n times, we obtain $\dim \mathcal{C}(B_n) \geq \dim \mathcal{C}(A)$, whence $\dim \mathcal{C}(B_n) > n$. This contradicts $\dim \mathcal{C}(B_n) = n$. \square

Remark 3.4

1. Every square matrix can be perturbed by arbitrarily small changes to a diagonalizable matrix with distinct eigenvalues. (The same is true of any collection of ASD matrices.) Proposition 3.3 says that, in general, not all these perturbations are 2-correctable.
2. The condition $\dim \mathcal{C}(A) > n$ fails only for nonderogatory matrices, which have $\dim \mathcal{C}(A) = n$.

The ϵ -perturbations used by Motzkin and Taussky in their proof of Theorem 2.1 are very special and of a different nature than the ones we shall use later on. In essence, the Motzkin-Taussky proof proceeds as follows. Suppose A_1 and A_2 are commuting $n \times n$ complex matrices which, by Proposition 3.1, we can assume are nilpotent. If A_1 is 1-regular, then A_2 is already a polynomial in A_1 , so we can perturb A_1 to a diagonalizable matrix whilst perturbing A_2 via the corresponding polynomial. Now suppose A_1 is not 1-regular. Then $\mathcal{C}(A_1)$ is decomposable and therefore we can choose a proper idempotent matrix $E \in \mathcal{C}(A_1)$. Now for any $\epsilon > 0$, the matrices A_1 and $A_2 + \epsilon E$ commute, and the latter has the two eigenvalues 0 and ϵ . We now have a proper splitting by Proposition 3.1, whence induction completes the proof. However, a similar technique cannot work for three commuting matrices A_1, A_2, A_3 by perturbing A_3 by ϵE for some proper idempotent $E \in \mathcal{C}(A_1, A_2)$, because $\mathcal{C}(A_1, A_2)$ can be indecomposable even when A_1 is not 1-regular.

Our final proposition of this section records an ϵ -perturbation of an arbitrary nonzero nilpotent matrix which introduces a new eigenvalue (of ϵ), but has the advantage over the Motzkin and Taussky type in that the perturbation is always 2-correctable. Moreover, an induction argument using this proposition and the splitting in Proposition 3.1 provides another proof of the Motzkin and Taussky result (Theorem 2.1).

Proposition 3.5. *Suppose J and K are commuting matrices with J nonzero and nilpotent. Let Q be a quasi-inverse for J (that is, $J = JQJ$ —if J is in Jordan form, one natural choice for Q is the transpose of J). Let $E = I - JQ$ and suppose $EQ^m = Q^m$ for some $m > 0$ (e.g. Q nilpotent). Let $\epsilon > 0$ and let $L = \epsilon Q + \epsilon^2 Q^2 + \dots + \epsilon^m Q^m$. Then*

- (1) *The matrices $\bar{J} = J + \epsilon E$ and $\bar{K} = K + LKE$ commute;*
- (2) *\bar{J} has 0 and ϵ as eigenvalues.*

Proof. Note the relations

$$(i) EJ = 0; \quad (ii) E^2 = E; \quad (iii) EK = EKE.$$

The third equation follows from $EK - EKE = EK(I - E) = EK(JQ) = E(KJ)Q = E(JK)Q = 0$ using (i). Now,

$$\begin{aligned}
\overline{J} \overline{K} &= JK + J L K E + \epsilon E K + \epsilon E L K E \\
&= JK + J \left(\sum_{i=1}^m \epsilon^i Q^i \right) K E + \epsilon E K + \epsilon E \left(\sum_{i=1}^m \epsilon^i Q^i \right) K E \\
&= JK + \epsilon (E K + J Q K E) + \sum_{i=2}^m \epsilon^i [J Q^i K E + E Q^{i-1} K E] \\
&\quad + \epsilon^{m+1} E Q^m K E. \quad (*)
\end{aligned}$$

Similarly,

$$\begin{aligned}
\overline{K} \overline{J} &= K J + L K E J + \epsilon K E + \epsilon L K E^2 \\
&= K J + \epsilon K E + \epsilon \left(\sum_{i=1}^m \epsilon^i Q^i \right) K E \quad \text{using (i) and (ii)} \\
&= K J + \epsilon K E + \sum_{i=2}^m \epsilon^i Q^{i-1} K E + \epsilon^{m+1} Q^m K E. \quad (**)
\end{aligned}$$

We now compare the expressions (*) and (**). We have $KJ = JK$ by assumption. Moreover, the coefficients of ϵ agree because $EK + JQKE = EK + (I - E)KE = EK + KE - EKE = KE$ by (iii). The ϵ^i terms agree for $i = 2, 3, \dots, m$ because $JQ^iKE + EQ^{i-1}KE = (I - E)Q^{i-1}KE + EQ^{i-1}KE = Q^{i-1}KE$. Finally, the ϵ^{m+1} terms agree because by assumption $EQ^m = Q^m$. Hence part (1) of the proposition holds.

We can see that ϵ is an eigenvalue of \overline{J} because $E[\epsilon I - (J + \epsilon E)] = \epsilon E - EJ - \epsilon E^2 = \epsilon E - 0 - \epsilon E = 0$ shows that $\epsilon I - \overline{J}$ is singular. Also, 0 is an eigenvalue of \overline{J} because if p is the nilpotent index of J , then $(J + \epsilon E)J^{p-1} = J^p + \epsilon EJ^{p-1} = 0 + 0 = 0$ which shows \overline{J} is singular. \square

4. The H-form

Determining which matrices commute with a given set of $n \times n$ commuting matrices is, in general, a difficult problem. Two tools appear to be helpful in tackling this problem: (1) a “standard form” for a given matrix and (2) restrictions on the form of the commuting matrices. In particular, upper triangular matrices are simpler to work with in deciding commuting relationships. Moreover, it is well known [8, Theorem 2.3.3] that a finite set of commuting matrices can be simultaneously upper triangularized. Unfortunately, the most well-known standard form, the Jordan form, is not compatible with retaining upper triangularity in commuting matrices. To help circumvent this problem, we define a new standard form, the H-form, for an $n \times n$ matrix over an algebraically closed field, that allows us to assume all commuting matrices are also upper triangular. (A bonus feature of our new form is that it allows a much simpler description of the centralizer of a matrix than does the Jordan form.) In

this section, we shall give three independent proofs for the existence of the H-form: (1) a simple “row operations” proof, (2) a derivation from the Jordan form, and (3) a module-theoretic proof. The last of these suggests that the H-form lives in a somewhat bigger universe than its Jordan counterpart, even though each can be derived from the other for matrices over an algebraically closed field.

Our basic H-matrices (defined below) can be viewed as blocked-matrix generalizations of a basic Jordan matrix

$$\begin{bmatrix} \lambda & 1 & & & & \\ & \lambda & 1 & & & \\ & & & \ddots & & \\ & & & & \ddots & \\ & & & & & \lambda & 1 \\ & & & & & & \lambda \end{bmatrix}$$

with associated eigenvalue λ , where we replace the λ 's by scalar matrices and the 1's by full column rank matrices in reduced row echelon form. Thus the diagonal blocks look like

$$\lambda I = \begin{bmatrix} \lambda & & & & \\ & \lambda & & & \\ & & \lambda & & \\ & & & \ddots & \\ & & & & \lambda \end{bmatrix}$$

for various identity matrices I that do not increase in size down the diagonal; and the first super-diagonal blocks look like

$$\begin{bmatrix} I \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & 1 \\ 0 & 0 & \cdots & & 0 \\ \vdots & & & & \vdots \end{bmatrix}$$

for rectangular matrices whose sizes are dictated by the diagonal block sizes. (We allow the possibility that there are no zero rows.) Unlike the Jordan form, our H-form does not allow multiple basic H-matrices for the same eigenvalue λ .

Before giving the formal definitions we note that in specifying the block structure of a blocked matrix, we need only specify the sizes of the (square) diagonal blocks (because the (i, j) block must be $n_i \times n_j$ where n_i and n_j are the i th and j th diagonal block sizes). Moreover, if the diagonal blocks have decreasing size, the whole block structure of an $n \times n$ matrix can be specified uniquely by a partition $n_1 + n_2 + \cdots + n_r = n$ of n with $n_1 \geq n_2 \geq \cdots \geq n_r \geq 1$.

Definition 4.1. A basic *H-matrix* with eigenvalue λ is an $n \times n$ matrix A of the following form: There is a partition $n_1 + n_2 + \cdots + n_r = n$ of n with $n_1 \geq n_2 \geq \cdots \geq n_r \geq 1$ such that when A is viewed as a blocked matrix with diagonal blocks of size n_1, n_2, \dots, n_r , the diagonal blocks are the $n_i \times n_i$ scalar matrices λI and the first super-diagonal blocks are full column rank $n_i \times n_{i+1}$ matrices in reduced row echelon form (i.e. an identity matrix followed by zero rows). All other blocks of A are zero. In this case, we say that A has an H-block structure (n_1, n_2, \dots, n_r) .

For example,

$$\left[\begin{array}{cc|cc|} \lambda & 0 & 1 & 0 & & \\ & \lambda & 0 & 1 & & \\ \hline & & \lambda & 0 & 1 & \\ & & & \lambda & 0 & \\ & & & & \lambda & 1 \\ & & & & & \lambda & 1 \\ & & & & & & \lambda \end{array} \right] \quad \text{and} \quad \left[\begin{array}{ccc|ccc} \lambda & 0 & 0 & 1 & 0 & 0 \\ & \lambda & 0 & 0 & 1 & 0 \\ & & \lambda & 0 & 0 & 1 \\ \hline & & & \lambda & 0 & 0 \\ & & & & \lambda & 0 \\ & & & & & \lambda \end{array} \right]$$

are legitimate basic H-matrices with block structures $(2, 2, 1, 1, 1)$ and $(3, 3)$ respectively. On the other hand, the matrix

$$\left[\begin{array}{ccccc} \lambda & 0 & 1 & & \\ & \lambda & 0 & 1 & \\ & & \lambda & 0 & 0 \\ & & & \lambda & 1 \\ & & & & \lambda & 1 \\ & & & & & \lambda & 1 \\ & & & & & & \lambda \end{array} \right]$$

is not a basic H-matrix because of the ill-positioned 1 in the 5th column. We can regard an $n \times n$ scalar matrix as a basic H-matrix with the trivial block structure (n) . At the other extreme, a basic H-matrix with block structure $(1, 1, 1, \dots, 1)$ is just a basic Jordan matrix.

Definition 4.2. Let A be a square matrix over an algebraically closed field F , and let $\lambda_1, \dots, \lambda_r$ be the distinct eigenvalues of A . We say that A is in *H-form* if A is a direct sum of basic H-matrices, one for each eigenvalue. In other words, A has the form

$$\left[\begin{array}{cccc} H_1 & & & \\ & H_2 & & \\ & & \ddots & \\ & & & H_r \end{array} \right],$$

where H_i is a basic H-matrix with eigenvalue λ_i for $i = 1, \dots, r$.

Remark 4.3

1. We will show in Theorem 4.7 and Proposition 4.8 that each square matrix is similar to a unique matrix in H-form.
2. A matrix J in Jordan form will be in H-form only in the case where for each eigenvalue, either there is just one basic Jordan block or all its basic Jordan blocks are 1×1 (so the matrix J is a direct sum of a diagonal matrix and a nonderogatory matrix). We again stress that our definition of H-form does not allow multiple basic H-blocks for the same eigenvalue.

We record the following simple observations concerning conjugations by elementary matrices, which will be used in the proof of our triangularization result, Theorem 4.7.

Lemma 4.4. *For $i \neq j$, let $E = E_{ij}(c)$ be the elementary matrix $I + ce_{ij}$. Then*

1. *Conjugating a matrix A by E (forming $E^{-1}AE$) has the effect of adding c times the i th column of A to its j th column, and then subtracting c times the j th row of the resulting matrix from its i th row.*
2. *If the i th column of A is zero, then the conjugation has the same effect on the i th row of A as the corresponding elementary row operation (of subtracting c times the j th row).*
3. *If the first d columns of A are zero, then any elementary row operation (including row swaps) on the first d rows of A can be realized as a conjugation by the corresponding elementary matrix.*

The centralizer of any basic $n \times n$ Jordan matrix J has a well-known and simple description as the set of upper triangular matrices K for which entries in the same super-diagonal (including the diagonal) are equal:

$$K = \begin{bmatrix} a & b & c & \cdots & & \\ & a & b & c & \cdots & \\ & & a & b & c & \cdots \\ & & & \ddots & & \\ & & & & \ddots & \\ & & & & & \ddots \end{bmatrix}.$$

That is, $K_{ij} = 0$ for $i > j$ and $K_{ij} = K_{i+1,j+1}$ for $1 \leq i \leq j \leq n-1$. (So the centralizer is the subalgebra generated by J .) For an $n \times n$ basic H-matrix A , the centralizer is a little more complicated but nevertheless has a similar description in terms of upper block-triangular matrices, if we weaken the requirement that the (i, j) and $(i+1, j+1)$ blocks be “equal” when the second block is strictly smaller.

Proposition 4.5. Let A be an $n \times n$ basic H -matrix with structure (n_1, \dots, n_r) where $r \geq 2$. Let K be an $n \times n$ matrix, blocked according to $n_i \times n_i$ diagonal blocks, and let K_{ij} denote its (i, j) block for $i, j = 1, \dots, r$. Then A and K commute if and only if K is an upper block-triangular matrix for which

$$K_{ij} = \begin{bmatrix} K_{i+1,j+1} & * \\ 0 & * \end{bmatrix} \quad \text{for } 1 \leq i \leq j \leq r-1,$$

where the column of asterisks disappears if $n_j = n_{j+1}$ and the $[0 \ *]$ row disappears if $n_i = n_{i+1}$.

Proof. By subtracting the diagonal of A , we can assume A is nilpotent. For $j = 2, \dots, r$ let I_j denote the $n_{j-1} \times n_j$ matrix having the $n_j \times n_j$ identity matrix as its upper part followed by $n_{j-1} - n_j$ zero rows. Then as a blocked matrix

$$A = \begin{bmatrix} 0 & I_2 & & & & \\ & 0 & I_3 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & 0 & I_r \\ & & & & & 0 \end{bmatrix}.$$

A simple calculation shows that K commutes with A precisely when $K_{21} = K_{31} = \dots = K_{r1} = 0$ and $K_{ij}I_{j+1} = I_{i+1}K_{i+1,j+1}$ for $1 \leq i, j \leq r-1$. The proposition now follows. \square

Remark 4.6. Using the upper block-triangular description in Proposition 4.5, and arguing inductively from the bottom row upwards on the number of free choices for entries in the i th row of blocks, one obtains the following simple formula for the dimension of the centralizer of a basic H -matrix A in terms of its H -block structure (n_1, n_2, \dots, n_r) :

$$\dim \mathcal{C}(A) = n_1^2 + n_2^2 + \dots + n_r^2.$$

This contrasts with the formula in terms of the Jordan structure (m_1, m_2, \dots, m_s) of A :

$$\dim \mathcal{C}(A) = m_1 + 3m_2 + 5m_3 + \dots + (2s-1)m_s.$$

Theorem 4.7. Let A_1, A_2, \dots, A_k be commuting $n \times n$ matrices over an algebraically closed field F . Then there is a similarity transformation which puts A_1 in H -form and simultaneously puts A_2, \dots, A_k in upper triangular form.

Proof. By Proposition 3.1 we can put A_1 in block diagonal form with each block having a single eigenvalue and different blocks having different eigenvalues. Because of commutativity, there is a matching block diagonal splitting of the other A_i but

without eigenvalue restrictions. Hence we can reduce to the case where A_1 has only a single eigenvalue, and, by subtraction of a scalar matrix, to the case where A_1 is a (nonzero) nilpotent matrix. We first put A_1 in H-form.

Let $d = \text{nullity } A_1$. After a similarity transformation (using a change of basis in which the first d members span the null space), we can assume that

$$A_1 = \begin{bmatrix} 0 & B \\ 0 & C \end{bmatrix},$$

where the matrix is blocked with a $d \times d$ top left hand corner. Moreover, by induction we can assume C is in H-form. Let $m = \text{nullity } C$ and note that m is the size of the first block in the H-block structure of C . Partition $B = [X, \bar{X}]$ where X is $d \times m$. Our matrix A_1 now looks like

$$\left[\begin{array}{c|c|c} 0 & X & \bar{X} \\ \hline & & 1 \\ & & \ddots \\ & & 1 \\ & & 0 \\ & & \vdots \\ 0 & & 0 & 1 \\ & & & \ddots \\ & & & 1 \\ & & & 0 \\ & & & \vdots \end{array} \right]$$

Since X must have full column rank, we have $d \geq m$. Observe that the submatrix of A_1 directly beneath \bar{X} has full column rank and echelon shape. Hence conjugating by various elementary matrices $E_{ij}(c)$ for $i = 1, \dots, d$ and $j = d + 1, \dots, n$ we can make $\bar{X} = 0$ by Lemma 4.4. Since the first d columns of A_1 are zero, by Lemma 4.4(3) X can be row reduced to

$$X = \left[\begin{array}{cccc} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ & & & 0 \\ & & & \vdots \\ & & & 0 \end{array} \right]$$

using conjugations by various elementary matrices, without affecting the other features of A_1 . Now A_1 is in H-form.

Thus we can assume A_1, A_2, \dots, A_k are commuting matrices with A_1 a basic H-matrix, say of structure (n_1, \dots, n_r) . By Proposition 4.5, each A_i is an upper block-triangular matrix with respect to this structure. We complete the proof by inductively constructing an invertible block-diagonal matrix $C = \text{diag}(C_1, \dots, C_r)$ which centralizes A_1 and conjugates A_2, \dots, A_k simultaneously to (properly) upper triangular matrices. We construct C_r, C_{r-1}, \dots, C_1 in this order.

The (r, r) blocks of A_2, \dots, A_k commute, so there is an invertible $r \times r$ matrix C_r which simultaneously conjugates these blocks to upper triangular matrices. Suppose we have constructed C_i for some $i > 1$. If $n_{i-1} = n_i$, we set $C_{i-1} = C_i$. Suppose $n_{i-1} > n_i$. Since A_2, \dots, A_k centralize A_1 , by Proposition 4.5 their $(i-1, i-1)$ blocks have the form

$$\begin{bmatrix} Y_j & * \\ 0 & Z_j \end{bmatrix} \quad \text{for } j = 2, \dots, k,$$

where Y_2, \dots, Y_k are their (i, i) blocks and Z_2, \dots, Z_k are $(n_{i-1} - n_i) \times (n_{i-1} - n_i)$ matrices. The Z_j commute because A_2, \dots, A_k commute. Choose an invertible $(n_{i-1} - n_i) \times (n_{i-1} - n_i)$ matrix D_{i-1} that simultaneously conjugates Z_2, \dots, Z_k to upper triangular matrices. Now set

$$C_{i-1} = \begin{bmatrix} C_i & 0 \\ 0 & D_{i-1} \end{bmatrix}.$$

It is clear that in this construction, each C_i simultaneously conjugates the $n_i \times n_i$ diagonal blocks of A_2, \dots, A_k to upper triangular matrices and moreover, by Proposition 4.5, $C = \text{diag}(C_1, \dots, C_r)$ centralizes A_1 . Therefore conjugation by C fixes A_1 and transforms A_2, \dots, A_k to upper triangular matrices, as desired. \square

To establish uniqueness of the H-form, we show that the H-block structure of a basic H-matrix A with eigenvalue λ is completely determined by the nullities of the powers of $A - \lambda I$. This is analogous to the situation for determining the basic Jordan matrices corresponding to λ (which give the Jordan structure for the eigenvalue λ), although for basic H-matrices the nullity connections are simpler.

Proposition 4.8. *If A is a basic H-matrix with eigenvalue λ and block structure (n_1, \dots, n_r) , then*

$$\begin{aligned} r &= \text{nilpotent index of } A - \lambda I, \\ n_1 &= \text{nullity } (A - \lambda I), \\ n_i &= \text{nullity } (A - \lambda I)^i - \text{nullity } (A - \lambda I)^{i-1} \quad \text{for } i = 2, \dots, r. \end{aligned}$$

Consequently, each square matrix is similar to a unique matrix in H-form (ignoring permutations of basic blocks).

Proof. Let $N = A - \lambda I$ and view N and its powers as blocked matrices with $n_i \times n_i$ diagonal blocks. Let I_i denote an appropriately sized matrix with n_i columns and having the $n_i \times n_i$ identity matrix as its upper part followed by zero rows. Then

$$N = \begin{bmatrix} 0 & I_2 & & & & \\ & 0 & I_3 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & 0 & I_r \\ & & & & & 0 \end{bmatrix} \quad \text{and}$$

$$N^i = \begin{bmatrix} 0 & \cdots & I_{i+1} & & & \\ & & & I_{i+2} & & \\ & & & & \ddots & \\ & & & & & I_r \\ & & & & & \\ & & & & & 0 \end{bmatrix}$$

for $i = 1, \dots, r-1$. Clearly N is nilpotent of index r . Now for $i = 1, \dots, r-1$ we have $\text{rank } N^i = n_{i+1} + n_{i+2} + \cdots + n_r$, giving $n_i = \text{rank } N^{i-1} - \text{rank } N^i = \text{nullity } N^i - \text{nullity } N^{i-1}$. Clearly this also holds for $i = r$. \square

We are indebted to Milen Yakimov for pointing out the following connection between the H-form and Jordan form.

Proposition 4.9. *The H and Jordan structures of any nilpotent $n \times n$ matrix A (more generally, a matrix with a single eigenvalue) are conjugate (“dual” or “transpose”) partitions of n . Moreover, the H-form and Jordan form of any square matrix are conjugate under a permutation transformation.*

Proof. We can assume that A is already a basic H-matrix, say with H-structure (n_1, n_2, \dots, n_r) . View A as the matrix of a transformation $T : F^n \rightarrow F^n$ relative to an ordered basis $\mathcal{B} = \{v_1, v_2, \dots, v_n\}$. Write $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_r$ where $\mathcal{B}_1 = \{v_1, \dots, v_{n_1}\}$ consists of the first n_1 basis vectors, \mathcal{B}_2 the next n_2 basis vectors, and so on. From the form of A , the action of T on \mathcal{B} is to annihilate \mathcal{B}_1 and then shift (in order) the n_i vectors in \mathcal{B}_i to the corresponding first n_i vectors in \mathcal{B}_{i-1} for $i = 2, \dots, r$. Now re-order the basis \mathcal{B} as $\mathcal{B}' = \mathcal{B}'_1 \cup \mathcal{B}'_2 \cup \cdots \cup \mathcal{B}'_s$ where \mathcal{B}'_1 consists of the first members of $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_r$ (in the ordering of \mathcal{B}), while \mathcal{B}'_2 consists of the second members of $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_r$ (no contribution from those \mathcal{B}_i with $|\mathcal{B}_i| = 1$), and so on down to \mathcal{B}'_s which consists of all the last members (of those \mathcal{B}_i with $|\mathcal{B}_i| = n_1$). We have the following Young diagram in which the boxes

contain the basis vectors of \mathcal{B} distributed in its rows, and the basis vectors of \mathcal{B}' distributed in its columns.

	\mathcal{B}'_1	\mathcal{B}'_2	\mathcal{B}'_3	\dots	\mathcal{B}'_s
\mathcal{B}_1				\dots	
\mathcal{B}_2			\dots		
\vdots					
\mathcal{B}_r		\dots			

Using our earlier observation on how T acts on vectors in \mathcal{B} , we see that T acts cyclically on each \mathcal{B}'_i , by shifting each vector to its predecessor and then annihilating the first. Thus the matrix J of T relative to \mathcal{B}' is the Jordan form of A , whence the Jordan structure of A is (m_1, m_2, \dots, m_s) where $m_i = |\mathcal{B}'_i|$ for $i = 1, 2, \dots, s$. Therefore, from the above diagram, the H and Jordan structures are conjugate partitions.

Our argument shows how to calculate the change of basis matrix C from \mathcal{B} to \mathcal{B}' , and hence obtain an explicit permutation matrix C with $J = CAC^{-1}$. For instance, suppose A has H-structure $(5, 3, 2)$. The Young diagram is then

v_1	v_2	v_3	v_4	v_5
v_6	v_7	v_8		
v_9	v_{10}			

so the Jordan structure is $(3, 3, 2, 1, 1)$ and the re-ordered basis is

$$\mathcal{B}' = \{v_1, v_6, v_9\} \cup \{v_2, v_7, v_{10}\} \cup \{v_3, v_8\} \cup \{v_4\} \cup \{v_5\}.$$

Hence C is the permutation matrix corresponding to the permutation $(2\ 4\ 9\ 3\ 7\ 5\ 10\ 6)$. \square

Remark 4.10

1. Proposition 4.9 provides an alternative path for establishing the existence and uniqueness of the H-form, by appealing to those same aspects of the Jordan form. Equally, our direct proofs in 4.7 and 4.8 for the H-form now yield a short “row operations” proof for the Jordan form.
2. Equating the two formulae in Remark 4.6 for the dimension of the centralizer of a basic H-matrix, and using the connection in Proposition 4.9, yields the number-theoretic relationship

$$n_1^2 + n_2^2 + \dots + n_r^2 = m_1 + 3m_2 + 5m_3 + \dots + (2s - 1)m_s$$

for any conjugate partitions (n_1, n_2, \dots, n_r) and (m_1, m_2, \dots, m_s) of n .

For future reference we record the following special case of Proposition 4.9 for 2-regular matrices.

Proposition 4.11. *Let J be a Jordan matrix with just two basic λ -blocks and Jordan structure (r, s) with $r \geq s$ (i.e. the block sizes are $r \times r$ and $s \times s$). Let A be the (basic) H-form of J . Then*

1. *If $r = s$, A has H-block structure (n_1, \dots, n_r) with $n_1 = n_2 = \dots = n_r = 2$. Thus, as a partitioned matrix,*

$$A - \lambda I = \begin{bmatrix} 0 & I_t \\ 0 & 0 \end{bmatrix},$$

where $t = 2s - 2$ is even. Here I_t denotes the $t \times t$ identity matrix, and 0 denotes a variable size zero matrix.

2. *If $r > s$, A has H-block structure (n_1, \dots, n_r) where $n_1 = n_2 = \dots = n_s = 2$ and $n_i = 1$ for $i = s + 1, \dots, r$. Thus, as a partitioned matrix,*

$$A - \lambda I = \begin{bmatrix} 0 & 0 & I_t & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_u \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

where $t = 2s - 1$ is odd and $u = r - s - 1$.

3. *Combined, the two statements say A has the form in (2) with*

$$t = 2s - 1 - \delta_{rs} \quad \text{and} \quad u = r - s - 1 + \delta_{rs}.$$

Example 4.12. The $(5, 3)$ Jordan matrix

$$\begin{bmatrix} \lambda & 1 & & & & & & \\ & \lambda & 1 & & & & & \\ & & \lambda & 1 & & & & \\ & & & \lambda & 1 & & & \\ & & & & \lambda & 0 & & \\ & & & & & \lambda & 1 & \\ & & & & & & \lambda & 1 \\ & & & & & & & \lambda \end{bmatrix}$$

has the H-form

$$\begin{bmatrix} \lambda & 0 & 1 & & & & & \\ & \lambda & 0 & 1 & & & & \\ & & \lambda & 0 & 1 & & & \\ & & & \lambda & 0 & 1 & & \\ & & & & \lambda & 0 & 1 & \\ & & & & & \lambda & 0 & 0 \\ & & & & & & \lambda & 1 \\ & & & & & & & \lambda \end{bmatrix}.$$

Consistent with Proposition 4.11, here we have $t = 2s - 1 = 2 \times 3 - 1 = 5$ and $u = r - s - 1 = 5 - 3 - 1 = 1$.

An H-form can also be established in module-theoretic terms for a nilpotent endomorphism in quite a general setting, as in our next proposition. This is essentially due to Goodearl [5] in the 1970's but with some recent modifications by Beidar et al. [3]. We first propose the following formulation of an H-form of a module endomorphism.

Definition 4.13. Suppose $\tau : P \rightarrow P$ is a nilpotent endomorphism of a (nonzero) projective module P over an arbitrary (and noncommutative) ring R . Then an H-form for τ is a direct sum decomposition

$$P = P_1 \oplus P_2 \oplus \cdots \oplus P_r$$

of P into nonzero submodules such that τ annihilates P_1 and maps P_i isomorphically onto a direct summand of P_{i-1} for $i = 2, \dots, r$.

Remark 4.14. In the broad setting of Definition 4.13, it is not clear what would be a natural formulation of a “Jordan form” for τ (given one wants uniqueness in the case of a linear transformation), without some additional assumptions about the nature of direct sum decompositions of P into cyclic indecomposables. (In general these do not exist even when R is a regular ring.) This may suggest that the concept of the H-form of a matrix over a field is a little more “basis-free” than its Jordan counterpart.

Proposition 4.15. Let $\tau : P \rightarrow P$ be a nilpotent endomorphism of a projective module P over a ring R . Then τ has an H-form precisely when all the powers τ^k of τ are (von Neumann) regular in the endomorphism ring $\text{End}_R(P)$. (Recall that an element a of a ring S is regular if $a = aba$ for some $b \in S$, equivalently, aS is a direct summand of S .)

Proof. In the case where R is a von Neumann regular ring and P is a finitely generated projective module, Goodearl in [5, Lemma 7.1] gave a decomposition similar to that in Definition 4.13 with τ mapping P_i onto P_{i-1} for $i = 2, \dots, r$ but not isomorphically. In [3, Lemma 3.5] it is shown that regularity of R in this result can be weakened to regularity of the powers of τ (and without insisting P be finitely generated), and then [3, Theorem 3.6] (and its proof) establish an H-form for τ as required by Definition 4.13. In fact, in the notation of the proof of [3, Theorem 3.6] (but working with P in place of R , τ in place of a , and r in place of n), an H-form $P = P_1 \oplus P_2 \oplus \cdots \oplus P_r$ for τ is provided by the “super-diagonals” of the scheme of the B_{ij} , that is,

$$P_i = \bigoplus_{j=1}^{r-i+1} B_{j,j+i-1}$$

for $i = 1, 2, \dots, r$. (The reader will notice some similarities in the proofs of [3, Theorem 3.6] and our Proposition 4.9.) \square

Of course, when $\tau : V \rightarrow V$ is a nilpotent linear transformation of a finite-dimensional vector space V over any field F , Proposition 4.15 applies (taking $R = F$ and

$P = V$) so τ has an H-form as a transformation. If the H-decomposition of V is $V = V_1 \oplus V_2 \oplus \cdots \oplus V_r$, one then obtains an H-form for a matrix of τ in a basis $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_r$ constructed as follows. Start with any basis \mathcal{B}_r for V_r . Next extend $\tau(\mathcal{B}_r)$ to a basis for V_{r-1} and call this \mathcal{B}_{r-1} . Continue in this way to inductively construct the \mathcal{B}_i for $i = r, r-1, \dots, 2$. Finally, take \mathcal{B}_1 to be any basis for V_1 (which is the null space of τ).

We now have three independent proofs for the existence of the H-form of a matrix over an algebraically closed field: (1) using “row operations” (4.7), (2) using the Jordan form (4.9), and (3) using a module decomposition (4.15).

5. Commuting matrices: the 2-regular case

In this section we determine the (nilpotent) matrices that commute with a given 2-regular nilpotent $n \times n$ matrix J in H-form where the parameter t is odd (see Proposition 4.11). Here again we work over an arbitrary algebraically closed field F . The $n \times n$ matrix with a 1 in the (i, j) position and 0's elsewhere is denoted e_{ij} . We maintain the following notation:

Notation 5.1. $J = \sum_{i=1}^t e_{i,i+2} + \sum_{i=t+2}^{n-1} e_{i,i+1}$ is a 2-regular nilpotent $n \times n$ matrix in H-form with t entries in its first identity block ($t \geq 3$). That is,

$$J = \begin{bmatrix} 0 & 0 & I_t & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_u \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

where the first two columns of zeroes in the partitioned matrix have width one, and the second and fourth rows of zeroes have depth one. (See Proposition 4.11.)

$K = (a_{ij})$ is a strictly upper triangular matrix which commutes with J .

In the 2-regular case, J is an H-matrix with block structure $(2, 2, \dots, 2, 1, 1, \dots, 1)$. Thus, by Proposition 4.5, K has the form

$$K = \begin{bmatrix} K_{2 \times 2} & K_{2 \times 1} \\ K_0 & K_{1 \times 1} \end{bmatrix},$$

where

$$K_{2 \times 2} = \begin{bmatrix} K_1 & K_2 & \cdots & K_{v-1} & K_v \\ 0 & K_1 & K_2 & \ddots & K_{v-1} \\ 0 & 0 & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & K_2 \\ 0 & 0 & 0 & 0 & K_1 \end{bmatrix}$$

is a blocked $v \times v$ matrix with 2×2 blocks and $v = (t+1)/2$,

$$K_{1 \times 1} = \begin{bmatrix} k_1 & k_2 & \cdots & k_{w-1} & k_w \\ 0 & k_1 & k_2 & \cdots & k_{w-1} \\ 0 & 0 & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & k_2 \\ 0 & 0 & 0 & 0 & k_1 \end{bmatrix}$$

is a $w \times w$ matrix with $w = u + 1$,

$$K_{2 \times 1} = \begin{bmatrix} K'_{v+1} & K'_{v+2} & \cdots & K'_{v+w} \\ \vdots & K'_{v+1} & K'_{v+2} & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ K'_3 & \ddots & K'_{w+1} & \vdots \\ K'_2 & K'_3 & \cdots & K'_{w+1} \end{bmatrix}$$

is a blocked $v \times w$ matrix with 2×1 blocks, and K_0 is a $w \times v$ zero matrix. Of course if the dimension t associated with the matrix J is even, then $K = K_{2 \times 2}$, and $K_0, K_{1 \times 1}, K_{2 \times 1}$ are empty. For the rest of this section, we will assume t is odd.

From Proposition 4.5, we have the following connections between blocks:

- 1st column of $K_i = \begin{bmatrix} k_i \\ 0 \end{bmatrix}$ whenever both K_i and k_i are defined;
- $K'_i = \begin{bmatrix} k_i \\ 0 \end{bmatrix}$ whenever both K'_i and k_i are defined;
- 1st column of $K_i = K'_i$ whenever both K'_i and K_i are defined.

The net result is a matrix of the form

$$K = \begin{bmatrix} K_1 & K_2 & K_3 & \cdots & K'_{v+1} & \cdots & \cdots & K'_{v+w} \\ & K_1 & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ & & \ddots & K_2 & K'_3 & \ddots & \ddots & \vdots \\ & & & K_1 & K'_2 & K'_3 & \ddots & \vdots \\ & & & & k_1 & k_2 & k_3 & \vdots \\ & & & & & k_1 & \ddots & \vdots \\ & & & & & & \ddots & k_2 \\ & & & & & & & k_1 \end{bmatrix}$$

satisfying the above relationships. We record one useful consequence of the form of $K = (a_{ij})$ in a lemma.

Lemma 5.2. With $K = (a_{ij})$ as in Notation 5.1, $a_{ij} = a_{i+2, j+2}$ for $i < j \leq t$.

The matrix J itself of course has the form of K , with $K_2 = I = 2 \times 2$ identity matrix, all other $K_i = 0$; $k_2 = 1$, all other $k_i = 0$; $K'_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and all other $K'_i = 0$.

If K is nilpotent, then K_1 is strictly upper triangular and $k_1 = 0$. By forming $K - k_2 J$, we can transform K to a matrix where $k_2 = 0$, $K'_2 = 0$, and the upper left corner of K_2 is zero. Similarly, $K - k_3 J^2$ produces a matrix where $k_3 = 0$, $K'_3 = 0$, and the upper left corner of K_3 is zero, assuming all these matrices are defined. It is possible, for example, that k_3 does not exist. In this case we would form $K - k'_3 J^2$, where k'_3 is the first row entry of K'_3 . The transformed version of K then has a 0 in the first row of K'_3 .

Continuing in this way, we can assume $k_i = 0$ for all i , the first row entry of K'_i is 0, and the upper left corner of K_i is 0 for all i . That is, $K = \begin{bmatrix} K_{2 \times 2} & K_{2 \times 1} \\ 0 & 0 \end{bmatrix}$ where the (i, j) entry of $K_{2 \times 2}$ is 0 if both i, j are odd, and the (i, j) entry of $K_{2 \times 1}$ is zero if i is odd. Such a K will be called *cleared out*. For the purposes of establishing the ASD property for commuting J, K, K', \dots (or the dimension of the subalgebra they generate), Lemma 2.4 entitles us to work with J and cleared out K, K', \dots because these generate the same subalgebra as J, K, K', \dots . We record two of the features of a cleared out matrix for future reference.

Lemma 5.3. *Let J be a matrix in H-form, with t odd. If $K = [a_{ij}]$ is a cleared out matrix commuting with J , then*

1. *If i, j are odd, then $a_{ij} = 0$.*
2. *If i is odd and $j \geq t + 2$, then $a_{ij} = 0$.*

6. Epsilon changes in 2-regular case: t odd

We return to working over the field of complex numbers \mathbb{C} . In the next two sections we will establish the following theorem.

Theorem 6.1. *The ASD property holds for three commuting matrices if one of them is 2-regular.*

Remark 6.2. Many of the techniques we use in the 2-regular case (particularly the use of the H-form and Theorem 4.7) have a wider applicability. For instance, they easily yield that any three commuting 5×5 matrices have the ASD property. We plan to present more general results in the future.

As mentioned in the Introduction, it is possible to deduce this theorem from the work of Neubauer and Sethuraman, specifically from [12, Theorem 15]. For it follows from this that commuting triples of matrices with the first 2-regular can be perturbed

to commuting ones for which the first matrix is 1-regular, and of course the latter are ASD because the second and third matrices will be polynomials in the first. To prove Theorem 6.1 by our purely matrix-theoretic methods it suffices, by Proposition 3.1 and Theorem 4.7, to establish the ASD property for commuting $n \times n$ complex matrices J, K, K' when J is a 2-regular nilpotent matrix in H-form and K, K' are strictly upper triangular matrices. In turn, we use the strategy discussed in Section 3 of perturbing J so as to introduce ϵ as a new eigenvalue.

Our arguments depend on whether $t \geq 1$ in Proposition 4.11 is even or odd. In both cases, we manage to make ϵ an eigenvalue of multiplicity one, but the degree of correctability of the perturbation varies according to whether t is even or odd. We handle the t odd case in this section and the t even case in the following section. If an ϵ -perturbation \bar{J} of J is 2-correctable, then by the argument used in the proof of Proposition 3.3, $\dim \mathcal{C}(\bar{J}) \geq \dim \mathcal{C}(J)$. Our perturbations applied repeatedly until the case $t = 0$ is reached can make the final perturbed \bar{J} diagonalizable with $n - 1$ distinct eigenvalues, whence $\dim \mathcal{C}(\bar{J}) = n + 2$. On the other hand, $\dim \mathcal{C}(J)$ can be as large as $2n$ (when t even), so it follows that not all our perturbations are 2-correctable. A closer analysis of our methods reveals that actually the perturbations used in the t odd case are k -correctable for all positive integers k , and the lack of correctability is confined to the t even case. There the perturbation of J has a “limited sort of 3-correctability within upper triangular matrices”. (This shows the usefulness of the H-form and the simultaneous triangularization Theorem 4.7 in these types of calculations.) In hindsight, the limited 3-correctability is about the best one could hope for in the t even case, in view of the four commuting upper triangular matrices in Example 2.7 ($t = 2$) failing the ASD property!

We retain the notation of 5.1 and also introduce the following matrices, including the proposed perturbations $\bar{J}, \bar{K}, \bar{K}'$ of our commuting J, K, K' in the case t is odd.

Definition 6.3

$$J = \begin{bmatrix} 0 & 0 & I_t & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_u \\ 0 & 0 & 0 & 0 \end{bmatrix} = \sum_{i=1}^t e_{i,i+2} + \sum_{i=t+2}^{n-1} e_{i,i+1},$$

$$Q = J^T,$$

$$E = e_{t+2,t+2},$$

$$\bar{J} = J + \epsilon E,$$

$$S = e_{11} + e_{33} + \cdots + e_{t+2,t+2},$$

$$\bar{K} = K - \epsilon K S Q,$$

$$\bar{K}' = K' - \epsilon K' S Q.$$

For the remainder of this section, we assume $t \geq 1$ is an odd integer. We also assume K and K' are cleared out, so the results of Section 5 hold. Our goal is to show that

the epsilon changes introduced in Definition 6.3 do not destroy commutativity. The first lemma establishes some basic relationships among our matrices. The equalities follow from direct computations.

Lemma 6.4. *In the notation of 6.3, if K and K' are cleared out then*

1. $SKS = 0$,
2. $EK = 0$,
3. $QJ = I - e_{11} - e_{22}$,
4. $Ke_{11} = 0$,
5. $SQ = SQS$,
6. $SQE = 0$,
7. $JSQ = S - E$.

Moreover, the same identities hold if K is replaced by K' .

Proof. (1) Multiplying on the left and right by S picks out odd rows and columns of K . These entries are zero by Lemma 5.3(1).

(2) Since EK has nonzero entries only in the $t + 2$ row, this matrix is zero by Lemma 5.3(2).

(3) Left multiplication by Q shifts rows 1 to t down two, while right multiplication by J shifts columns 1 to t two to the right.

(4) Since K is cleared out and upper triangular, the first column is zero. See Lemma 5.3(1).

(5,6) The matrix SQ has entries b_{ij} that are nonzero only if i, j are odd with $i = j + 2 \leq t + 2$. It is immediate that $SQS = SQ$ and $SQE = 0$.

(7) Multiplication on the left by J shifts rows up two, while multiplication by Q on the right shifts columns two to the left.

The last sentence of the lemma is clear from the identical form for K and K' . \square

Proposition 6.5. *In the notation of 6.3, if t is odd and K, K' are cleared out, then $\overline{J} \overline{K} = \overline{K} \overline{J}$, and $\overline{J} \overline{K}' = \overline{K}' \overline{J}$.*

Proof. It suffices to prove only the first equality. First, by definition,

$$\overline{J} \overline{K} = JK - \epsilon JKSQ + \epsilon EK - \epsilon^2 EKSQ.$$

We can use (2) of Lemma 6.4 to eliminate ϵEK and $\epsilon^2 EKSQ$. Then

$$\begin{aligned} \overline{J} \overline{K} &= JK - \epsilon JKSQ = JK - \epsilon KJSQ = JK - \epsilon K(S - E) \\ &= JK - \epsilon KS + \epsilon KE, \end{aligned}$$

by (7). Next

$$\overline{K} \overline{J} = KJ - \epsilon KJSQ + \epsilon KE - \epsilon^2 KSQE.$$

Finally we use (3) and (6) of Lemma 6.4 to write

$$\overline{K} \overline{J} = KJ - \epsilon KS(I - e_{11} - e_{22}) + \epsilon KE = KJ - \epsilon KS + \epsilon KE. \quad \square$$

Proposition 6.6. *In the notation of 6.3, if t is odd and J, K, K' commute, with K and K' cleared out, then $\overline{K} \overline{K}' = \overline{K}' \overline{K}$.*

Proof. By 6.3,

$$\overline{K} \overline{K}' = KK' - \epsilon KK'SQ - \epsilon KSQK' + \epsilon^2 KSQK'SQ.$$

The expression for $\overline{K}' \overline{K}$ is obtained by interchanging K and K' . We have $KSQK'S = KSQSK'S = 0$ by (5) and (1) of Lemma 6.4. Thus, it remains to show

$$(*) \quad KSQK' = K'SQK.$$

Let $W = KSQK'$. We will show that the matrix W has the following properties:

- (1) W has columns $t+3, t+4, \dots, n$ all zero.
- (2) Let $F = e_{11} + e_{22} + \dots + e_{t+2, t+2}$. Then $W = (FKF)S(FQF)(FK'F)$.
That is, W is the product of the top left $(t+2) \times (t+2)$ corners of K, S, Q, K' .

For (1), note that by Lemma 5.3(2), K' has nonzero entries only in the even rows of columns $t+2, t+3, \dots, n$. Hence SK' has columns $t+2, \dots, n$ all zero, so that $W = (KSQ)(SK')$ does as well.

To show (2), note by (1) that $W = WF$. Also $KS = FKS$ by the upper triangularity of K , so that $W = FW$. Again using triangularity, $K'F = FK'F$. Finally, $S = FSF$ is clear. Thus,

$$W = FWF = FK(FSF)Q(FK'F) = (FKF)S(FQF)(FK'F).$$

A final lemma (taking $m = t+2$) completes the proof of (*). \square

Lemma 6.7. *Let m be an odd positive integer. Let S and Q be the $m \times m$ matrices*

$$S = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} = \text{diag}(1, 0, 1, 0, \dots, 1),$$

$$Q = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 \end{bmatrix}.$$

Suppose that K and K' are nilpotent upper triangular commuting $m \times m$ matrices satisfying

- (i) $a_{ij} = 0$ when both i and j are odd;
- (ii) $a_{ij} = a_{i+2,j+2}$ for $i < j \leq m-2$.

Then $KSQK' = K'SQK$.

Proof. Let

$$R = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & & \vdots \\ \vdots & & & & & \vdots \\ 0 & 0 & \cdots & 0 & 1 & 0 \end{bmatrix}$$

and note that $R^2 = Q$. From (i) and (ii) of the hypotheses, KS has the form

$$KS = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & a & 0 & b & 0 & c & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & a & 0 & b & \cdots & 0 \\ \vdots & & & & & & & & \vdots \\ \vdots & & & & & & & & \vdots \\ \vdots & & & & & & & & \vdots \\ 0 & 0 & 0 & \cdots & \cdots & 0 & 0 & a \\ 0 & 0 & 0 & \cdots & \cdots & 0 & 0 & 0 \end{bmatrix}$$

so that

$$KSR = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & a & 0 & b & 0 & c & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & a & 0 & b & 0 & \cdots & 0 \\ \vdots & & & & & & & & \vdots \\ \vdots & & & & & & & & \vdots \\ \vdots & & & & & & & & \vdots \\ \vdots & & & & & & & & \vdots \\ 0 & 0 & 0 & \cdots & \cdots & 0 & a & 0 \\ 0 & 0 & 0 & \cdots & \cdots & 0 & 0 & 0 \end{bmatrix}.$$

Similarly,

$$RSK' = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & a' & 0 & b' & 0 & c' & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & a' & 0 & b' & 0 & \cdots & 0 \\ \vdots & & & & & & & & \vdots \\ \vdots & & & & & & & & \vdots \\ \vdots & & & & & & & & \vdots \\ \vdots & & & & & & & & \vdots \\ 0 & 0 & 0 & \cdots & \cdots & 0 & a' & 0 \\ 0 & 0 & 0 & \cdots & \cdots & 0 & 0 & 0 \end{bmatrix}.$$

The map that deletes the zero odd columns and the zero odd rows of the algebra of matrices having the form of KSR and RSK' is an algebra isomorphism. Under this map the images of the two matrices have the form

$$\begin{bmatrix} u & v & w & x & y & \cdots \\ 0 & u & v & w & x & \cdots \\ 0 & 0 & u & v & w & \cdots \\ \vdots & & & & & \\ \vdots & & & & & \\ 0 & 0 & 0 & \cdots & u \end{bmatrix}.$$

It is well known that such matrices commute (they are just polynomials in some basic Jordan matrix). Therefore KSR and RSK' commute whence, using Lemma 6.4(5), we obtain

$$\begin{aligned} KSQK' &= KSQSK' = KSR^2SK' \\ &= (KSR)(RSK') = (RSK')(KSR) \end{aligned}$$

$$\begin{aligned}
 &= RS(K'K)SR = RS(KK')SR \\
 &= \dots \\
 &= K'SQK.
 \end{aligned}$$

This completes the proof of the lemma and therefore of the Proposition 6.6. \square

We remark that the arguments in this section can be applied to any number of commuting matrices J, K, K', K'', \dots

7. Epsilon changes in the 2-regular case: t even

We proceed to the case when t is even ($t \geq 2$). In this case, $t = n - 2$ by Proposition 4.11. As above, J denotes the $n \times n$ nilpotent matrix in H-form, with n even and $t = n - 2$: $J = \begin{bmatrix} 0 & I_t \\ 0 & 0 \end{bmatrix} = \sum_{i=1}^t e_{i,i+2}$. Then $Q = J^{tr} = \sum_{i=1}^t e_{i+2,i}$. Set $m = n/2$.

When n is even, we can divide our matrices into 2×2 blocks to simplify notation and calculations. By Theorem 4.7 and Proposition 4.5, we may assume that any matrix commuting with J has the following form:

$$K = \begin{bmatrix} D_0 & D_1 & D_2 & \cdots & D_{m-2} & D_{m-1} \\ \mathbf{0} & D_0 & D_1 & D_2 & \ddots & D_{m-2} \\ \mathbf{0} & \mathbf{0} & D_0 & D_1 & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ddots & \ddots & D_2 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & D_0 & D_1 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & D_0 \end{bmatrix},$$

where the D_i are 2×2 matrices, D_0 is upper triangular, and $\mathbf{0}$ is the 2×2 zero matrix. For a 2×2 matrix D , we will use the notation $[D]$ to denote the $n \times n$ matrix with D 's down the main diagonal. Then K can be written uniquely as

$$K = [D_0] + [D_1]J + [D_2]J^2 + \cdots + [D_{m-1}]J^{m-1}.$$

The matrices we are working with are also nilpotent, whence D_0 must be strictly upper triangular.

If we are given a list of matrices commuting with J (and with each other), then we choose a matrix $K = [D_0] + [D_1]J + \cdots + [D_{m-1}]J^{m-1}$ in the list such that its first index h for which D_h is not a scalar matrix is minimal among all such indices over all the matrices in our list. (We can assume such an index exists, otherwise the algebra generated by J and these matrices will be generated by J alone, in which case the ASD property is automatic.)

By conjugating K (and all other commuting matrices) by a block diagonal matrix $[P]$, P a 2×2 invertible matrix, we may assume $D_h = \begin{bmatrix} a & b \\ 0 & c \end{bmatrix}$ is in Jordan form.

(This conjugation leaves J unchanged.) Take $K - cJ^h$ to change the $(2, 2)$ entry of D_h to 0. Then conjugate again to put D_h back in Jordan form. There are two possibilities (after multiplying K by a suitable scalar): $D_h = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ (Case (i)) and $D_h = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ (Case (ii)). Suppose now that K' is another on our list of matrices commuting with both J and K . Because of our choice of h we may assume (after subtracting scalar multiples of I, J, \dots, J^{h-1}) that K' has the form $K' = [D'_h]J^h + \dots + [D'_{m-1}]J^{m-1}$. We can use the powers J^i and the products KJ^i to successively clear out the D'_i so that they have the form $D'_i = \begin{bmatrix} c'_i & 0 \\ d'_i & 0 \end{bmatrix}$ in Case (i) and $D'_i = \begin{bmatrix} 0 & c'_i \\ d'_i & 0 \end{bmatrix}$ in Case (ii), for all $i \geq h$.

Assume first that $2h < m$. Then $KK' = K'K$ implies $D_h D'_h = D'_h D_h$. In either Case (i) or Case (ii) the conclusion is $c'_h = d'_h = 0$. Thus, $K' = [D'_{h+1}]J^{h+1} + \dots + [D'_{m-1}]J^{m-1}$. If $2h + 1 < m$, then $D_h D'_{h+1} = D'_{h+1} D_h$ implies $c'_{h+1} = d'_{h+1} = 0$. Continuing this argument gives $c'_i = d'_i = 0$ for $h + i < m$. It follows that K' may be assumed to have the form $K' = [D'_{m-h}]J^{m-h} + \dots + [D'_{m-1}]J^{m-1}$. If $2h \geq m$, then we cannot conclude that any extra $D'_i = 0$ so the form of K' cannot be simplified.

As a result of the discussion above, we can assume the commuting matrices K, K' have the following form: $K = [D_h]J^h + \dots + [D_{m-1}]J^{m-1}$, $K' = [D'_g]J^g + \dots + [D'_{m-1}]J^{m-1}$, where $g \geq h$ and $g + h \geq m$. Since our arguments below no longer require any special properties of D_h (it can even be zero), we next make the simplifying assumption $g = m - h$. This is possible because if $h \geq m/2$, we can change h to the greatest integer in $m/2$ and set $D_h = 0$ if the original h was strictly greater than $m/2$; while if $h < m/2$, we have shown $g = m - h$. Thus, we have a

Standard Form: With $h \leq m/2 = n/4$,

$$\begin{aligned} K &= [D_h]J^h + \dots + [D_{m-1}]J^{m-1}, \\ K' &= [D'_{m-h}]J^{m-h} + \dots + [D'_{m-1}]J^{m-1}. \end{aligned}$$

Remark. This form shows that if K has its D_0 a nonscalar matrix ($h = 0$), then any matrix K' commuting with J and K is already in the algebra they generate.

We now consider two cases:

- (1) D_h is diagonalizable.
- (2) D_h is not diagonalizable.

Case (1): We can conjugate J, K, K' by a block diagonal matrix with a 2×2 matrix P along the diagonal so that the new D_h is diagonal. Next subtract a scalar multiple of J^h so that D_h has the form $D_h = \begin{bmatrix} * & 0 \\ 0 & 0 \end{bmatrix}$. We introduce epsilon changes to J, K, K' as follows:

Notation 7.1. Case (1).

$$K = [D_h]J^h + \cdots + [D_{m-1}]J^{m-1} \quad \text{with } D_h = \begin{bmatrix} * & 0 \\ 0 & 0 \end{bmatrix},$$

$$K' = [D'_{m-h}]J^{m-h} + \cdots + [D'_{m-1}]J^{m-1},$$

$$E = e_{22},$$

$$\bar{J} = J + \epsilon E,$$

$$T = e_{22} + e_{44} + \cdots + e_{n-2, n-2},$$

$$\bar{K} = K - \epsilon QTK,$$

$$\bar{K}' = K' - \epsilon QTK'.$$

A technical lemma establishes useful elementary relationships among the matrices defined above.

Lemma 7.2. With notation as in 7.1

1. $JQT = T$,
2. $Q TJ = -E + T + e_{nn}$,
3. $EQ = 0$,
4. $KE = 0 = K'E$,
5. $e_{nn}K = 0 = e_{nn}K'$,
6. $0 = KK' = K'K = K'QTK = KQTK'$.

Proof. Parts 1 and 2 are straightforward applications of the definitions.

(3) $EQ = e_{22}Q = 0$ since Q has zero second row.

(4) $Ke_{22} = 0 = K'e_{22}$, since K and K' have zero second column.

(5) $e_{nn}K = 0$ because K is strictly upper triangular.

(6) From the standard form above, together with the fact that $J^m = 0$, we have $KK' = 0 = K'K$. Note that K begins with $2h$ columns of zeroes while K' ends with $n - 2h$ rows of zeroes. Now right multiplication by QT shifts the even numbered columns of K two to the left. However, the $2h + 2$ column of K is zero because of the form of D_h , whence the first $2h$ columns of KQT are also zero. Hence $KQTK' = 0$.

Similarly, K' begins with $n - 2h$ columns of zeroes and K ends with $2h + 1$ rows of zeroes. Next note that left multiplication by QT shifts the even numbered rows of K down two. But the $n - 2h$ row of K is zero because of the form of D_h . Thus, QTK still ends with $2h$ rows of zeroes and so $K'QTK = 0$. \square

Proposition 7.3. For Case (1), with the notation of 7.1 we have $\bar{J}\bar{K} = \bar{K}\bar{J}$, $\bar{J}\bar{K}' = \bar{K}'\bar{J}$ and $\bar{K}\bar{K}' = \bar{K}'\bar{K}$.

Proof. Using the definitions and identities from Lemma 7.2,

$$\overline{J} \overline{K} = JK - \epsilon J Q T K + \epsilon E K - \epsilon^2 E Q T K = JK - \epsilon T K + \epsilon E K - 0.$$

On the other hand,

$$\overline{K} \overline{J} = K J + \epsilon K E - \epsilon Q T K J - \epsilon^2 Q T K E.$$

Again using identities in Lemma 7.2, along with $Q T K J = Q T J K$,

$$\overline{K} \overline{J} = K J + 0 - \epsilon(-E + T + e_{nn})K - 0.$$

After noting $e_{nn}K = 0$, we see that the expressions for $\overline{J} \overline{K}$ and $\overline{K} \overline{J}$ are equal. To show $\overline{J} \overline{K}' = \overline{K}' \overline{J}'$, the argument is the same.

Finally,

$$\overline{K} \overline{K}' = K K' - \epsilon K Q T K' - \epsilon Q T K K' + \epsilon^2 Q T K Q T K' = K K'$$

using the identities in (6). Similarly,

$$\overline{K}' \overline{K} = K' K - \epsilon K' Q T K - \epsilon Q T K' K + \epsilon^2 Q T K' Q T K = K' K$$

using the identities in (6). Since K and K' commute, the proof is complete. \square

Remark. It is easy to check for Case (1) that if K'' is another commuting matrix with the same form as K' , then \overline{K}' and \overline{K}'' commute whenever the 2×2 block matrices in their expansions satisfy $D'_i = D''_i = 0$ for $i \leq n/2$. Thus, we can introduce epsilon changes to any number of commuting matrices if the beginning indices g in the expansion of the matrices other than J and K satisfy $g > n/2$.

Case (2): D_h is not diagonalizable. Here we conjugate as before with a block diagonal matrix to put D_h into Jordan form. By subtracting a scalar multiple of J^h from K we may assume $D_h = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. Let L be the block diagonal matrix with repeated 2×2 blocks $\begin{bmatrix} 1 & 0 \\ \epsilon & 1 \end{bmatrix}$. Then L centralizes J (by Proposition 4.5). Moreover, the matrices $J, KL, K'L$ commute. In fact $KLK'L = 0 = K'CLKL$ by the same arguments that show $KK' = 0 = K'K$. But KL has for its “coefficient D_h ” the matrix $\begin{bmatrix} \epsilon & 1 \\ 0 & 0 \end{bmatrix}$ which is diagonalizable. Therefore by Case (1) we can make epsilon changes to $J, KL, K'L$ which will yield the desired epsilon changes to J, K, K' .

We can now complete the proof of Theorem 6.1.

Proof of Theorem 6.1. By induction we can assume that $n \geq 2$ and that the theorem holds for matrices of size smaller than $n \times n$. Let J, K, K' be commuting $n \times n$ matrices where J is 2-regular. To establish the ASD property for these matrices, by Proposition 3.1 and Theorem 4.7 we may assume that J is nilpotent in H-form and K, K' are strictly upper triangular. Let $\epsilon > 0$ be given. By our arguments in this and the previous section, namely the confluence of Propositions 6.5, 6.6 and 7.3, together with the reduction of Case(2) to Case(1), we can obtain ϵ -perturbations $\overline{J}, \overline{K}, \overline{K}'$

of J, K, K' which remain commuting but where \bar{J} is a 2-regular matrix with two distinct eigenvalues (0 and ϵ). There is now a nontrivial simultaneous block diagonal splitting of $\bar{J}, \bar{K}, \bar{K}'$, courtesy of Proposition 3.1. On each of its blocks, \bar{J} will be (at most) 2-regular. Thus by induction, corresponding blocks of $\bar{J}, \bar{K}, \bar{K}'$ have the ASD property and therefore so too do their parents by Proposition 3.1. In turn, of course, this shows that J, K, K' are ASD, as desired. \square

8. Bounds on $\dim \mathbb{C}[A_1, \dots, A_k]$

As a corollary to Theorems 6.1 and 2.5, we obtain the following result of Neubauer and Sethuraman [12, Theorem 15].

Theorem 8.1 (Neubauer and Sethuraman). *If A_1, A_2, A_3 are commuting $n \times n$ matrices and at least one is 2-regular, then $\dim \mathbb{C}[A_1, A_2, A_3] \leq n$.*

Proof. The three matrices have the ASD property by Theorem 6.1, whence $\dim \mathbb{C}[A_1, A_2, A_3] \leq n$ by Theorem 2.5. \square

Example 2.7 shows that the ASD property can fail for more than three commuting matrices even when one of them is 2-regular. So in that case we cannot use our argument in Theorem 8.1 to bound the dimension of the subalgebra such matrices generate. Our techniques, however, still yield the following (sharp) upper bound.

Theorem 8.2. *Let A_1, \dots, A_k be commuting $n \times n$ matrices over the complex numbers, at least one of which is 2-regular. Then $\dim \mathbb{C}[A_1, \dots, A_k] \leq 5n/4$.*

Proof. Let $\mathcal{A} = \mathbb{C}[A_1, \dots, A_k]$ where A_1, \dots, A_k are commuting $n \times n$ matrices with A_1 a 2-regular matrix. We can clearly assume that $\{I, A_1, \dots, A_k\}$ is a vector space basis for \mathcal{A} . By Proposition 3.1 and Theorem 4.7 we can also assume that A_1, \dots, A_k are strictly upper triangular and that A_1 is in H-form. Let $J = A_1$. Thus J has the form

$$J = \begin{bmatrix} 0 & 0 & I_t & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_u \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

We consider two cases.

Case 1: t is even.

In this case $u = 0$, $t = n - 2$ and, as shown in Section 7, the general element of \mathcal{A} can be written as

$$(*) \quad [D_0]J^0 + [D_1]J + \dots + [D_{m-1}]J^{m-1},$$

where $m = n/2$ and each $[D_i]$ is a block diagonal matrix with repeated 2×2 diagonal blocks D_i . There is no loss of generality in assuming that \mathcal{A} is not generated as an algebra by J alone (because then we would know $\dim \mathcal{A} \leq n$), whence some A_i has a coefficient in (*) which is not a scalar matrix. By the Standard Form preceding Notation 7.1, there exist an integer h with $1 \leq h \leq m/2$ and a matrix $K \in \mathcal{A}$ of the form

$$K = [D_h]J^h + [D_{h+1}]J^{h+1} + \cdots + [D_{m-1}]J^{m-1}$$

such that, as a vector space, \mathcal{A} is spanned by $J^0, J, J^2, \dots, J^{m-1}, K, KJ, KJ^2, \dots, KJ^{m-h-1}$ and various matrices of the form $[D'_{m-h}]J^{m-h} + \cdots + [D'_{m-1}]J^{m-1}$. Therefore, on noting the form of K , we see that \mathcal{A} is spanned by

$$J^0, J, \dots, J^{m-h-1}, K, KJ, \dots, KJ^{m-2h-1}$$

and various matrices of the form

$$[D'_{m-h}]J^{m-h} + \cdots + [D'_{m-1}]J^{m-1}.$$

The first group has $(m-h) + (m-2h) = 2m-3h$ members. The second group clearly span a vector space of dimension at most $4h$. Therefore

$$\dim \mathcal{A} \leq 2m-3h+4h = 2m+h \leq 2m+m/2 = 5m/2 = 5n/4$$

as desired.

Case 2: t is odd.

In this case, by Section 6, we can perturb A_1, A_2, \dots, A_k to commuting matrices $\bar{A}_1, \bar{A}_2, \dots, \bar{A}_k$ such that \bar{A}_1 is 2-regular with two distinct eigenvalues. By Proposition 2.3, we can choose the perturbations small enough to ensure $I, \bar{A}_1, \dots, \bar{A}_k$ are linearly independent. Now by Proposition 3.1 we have, after a similarity transformation, nontrivial splittings

$$\bar{A}_i = \text{diag}(B_{i1}, B_{i2}, \dots, B_{ir}) \quad \text{for } i = 1, \dots, k,$$

where each B_{ij} is an $n_j \times n_j$ matrix and each B_{1j} is 2-regular. For fixed j , the matrices $B_{1j}, B_{2j}, \dots, B_{kj}$ commute, whence by induction (or repeated splittings to the t even case) we have $\dim \mathbb{C}[B_{1j}, B_{2j}, \dots, B_{kj}] \leq 5n_j/4$. Hence

$$\begin{aligned} \dim \mathbb{C}[A_1, \dots, A_k] &\leq \dim \mathbb{C}[\bar{A}_1, \dots, \bar{A}_k] \\ &\leq \sum_{j=1}^r \dim \mathbb{C}[B_{1j}, \dots, B_{kj}] \\ &\leq \sum_{j=1}^r 5n_j/4 \\ &= 5n/4 \end{aligned}$$

which completes the proof. \square

The following example shows that the $5n/4$ bound in Theorem 8.2 is sharp.

Example 8.3. For each positive integer n which is a multiple of 4, there is a commutative subalgebra \mathcal{A} of complex $n \times n$ matrices containing a 2-regular matrix and having $\dim \mathcal{A} = 5n/4$.

Proof. Suppose $n = 4h$. Let J be the basic nilpotent $n \times n$ H-matrix with H-structure $(2, 2, \dots, 2)$, that is, as a blocked matrix with 2×2 blocks

$$J = \begin{bmatrix} 0 & I & & & & \\ & 0 & I & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & 0 & I \\ & & & & & 0 \end{bmatrix}.$$

Let \mathcal{A} be the subalgebra of all matrices of the form

$$D_0 J^0 + D_1 J + D_2 J^2 + \dots + D_{2h-1} J^{2h-1},$$

where each D_i is a block diagonal matrix with repeated 2×2 diagonal blocks but with the restriction that D_0, D_1, \dots, D_{h-1} must be scalar matrices. Note that these matrices centralize J (Proposition 4.5), and the product of any pair with $D_0 = D_1 = \dots = D_{h-1} = 0$ results in zero. Thus \mathcal{A} is commutative, it contains the 2-regular matrix J , and has

$$\dim \mathcal{A} = h + 4h = 5n/4. \quad \square$$

9. Open questions

We conclude with a list of open problems arising from our work.

1. What happens with the ASD property for commuting (complex) matrices (even just three) when one of them is d -regular, $d \geq 3$?
2. If commuting $n \times n$ matrices A_1, A_2, \dots, A_k generate a subalgebra of dimension at most n , must A_1, A_2, \dots, A_k have the ASD property (cf. Theorem 2.5)?
3. Suppose A_1, A_2, \dots, A_k are commuting $n \times n$ matrices with A_1 d -regular. Find a sharp upper bound, in terms of d and n , on $\dim \mathbb{C}[A_1, A_2, \dots, A_k]$. In particular, what happens for $d = 3$? (The answers for $d = 1$ and 2 are respectively n and $5n/4$.)

Acknowledgments

We are grateful to Mike Steel for bringing the ASD problem to our attention. We also thank the Universities of Canterbury and Connecticut for their hospitality while this research was carried out.

References

- [1] E.S. Allman, J.A. Rhodes, Phylogenetic invariants for the general Markov model of sequence mutation, *Math. Biosci.* 186 (2003) 133–144.
- [2] J. Barria, P. Halmos, Vector bases for two commuting matrices, *Linear and Multilinear Algebra* 27 (1990) 147–157.
- [3] K.I. Beidar, K.C. O'Meara, R.M. Raphael, On uniform diagonalisation of matrices over regular rings and one-accessible regular algebras, *Comm. Algebra* 32 (2004) 3543–3562.
- [4] M. Gerstenhaber, On dominance and varieties of commuting matrices, *Ann. Math.* 73 (1961) 324–348.
- [5] K.R. Goodearl, *Von Neumann Regular Rings*, Pitman, London, 1979 (second ed., Krieger, Malabar, FL, 1991).
- [6] Robert M. Guralnick, A note on commuting pairs of matrices, *Linear and Multilinear Algebra* 31 (1992) 71–75.
- [7] Robert M. Guralnick, B.A. Sethuraman, Commuting pairs and triples of matrices and related varieties, *Linear Algebra Appl.* 310 (2000) 139–148.
- [8] Roger A. Horn, Charles R. Johnson, *Matrix Analysis*, Cambridge University Press, New York, 1985.
- [9] T. Laffey, S. Lazarus, Two-generated commutative matrix subalgebras, *Linear Algebra Appl.* 147 (1991) 249–273.
- [10] T. Motzkin, O. Taussky, Pairs of matrices with property L II, *Trans. Amer. Math. Soc.* 80 (1955) 387–401.
- [11] M.J. Neubauer, D.J. Saltman, Two-generated commutative subalgebras of $M_n(F)$, *J. Algebra* 164 (1994) 545–562.
- [12] M.J. Neubauer, B.A. Sethuraman, Commuting pairs in the centralizers of two-regular matrices, *J. Algebra* 214 (1999) 174–181.
- [13] A.R. Wadsworth, On commuting pairs of matrices, *Linear and Multilinear Algebra* 27 (1990) 159–162.